

Second Order Asymptotics for Quantum Hypothesis Testing

Ke Li*

Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543
(Dated: August 8, 2012)

In the asymptotic theory of quantum hypothesis testing, the error probability of the first kind jumps sharply from zero to one when the error exponent of the second kind passes by the point of the relative entropy of the two states, in an increasing way. This is well known as the direct part and strong converse of quantum Stein's lemma. Here we look into the behavior of this sudden change and have make it clear how the error of first kind grows according to a lower order of the error exponent of the second kind, and hence, we obtain the second order asymptotics for quantum hypothesis testing. Our method is elementary, based on basic linear algebra and probability theory. It deals with the achievability part and the converse part in a unified framework, with a clear geometric picture.

Hypothesis testing is an important subject in mathematical statistics and information theory. A typical scenario which is of fundamental significance is the asymptotic hypothesis testing problem with two hypotheses, each being many copies of independent and identically-distributed instances, occurring according to some given statistical description ρ or σ , respectively. Here the statistical descriptions ρ and σ are probability distributions in classical setting and quantum states which are positive semi-definite matrices with trace 1 in quantum mechanics. The task is to minimize the error probabilities of mistaking one hypothesis for the other under certain figure of merit.

Classically, this problem has been well understood [1–6]. Moving to the quantum case, it becomes much more difficult due to the non-commutativity of the quantum states ρ and σ , and the more complicated mechanics for observing the physical systems of interest (i.e., quantum measurement). Substantial achievements have already been made in the asymptotic theory of quantum hypothesis testing. Most notably, these include the establishing of the quantum Stein's lemma with a strong converse [7, 8], the quantum Chernoff bound [9, 10], and the quantum Hoeffding bound [11–13]. Some other important progress can be found in [13–19].

The Second order asymptotics is another interesting topic in mathematical statistics [20] and information theory [21–26]. In asymptotic analysis, not only does the second order term improve the accuracy of solution to the problem of concern, but also it plays a dominative role at points where the quantity of concern is asymptotically discontinuous as a function of the first order of some other quantity. This is exactly the case for capacity of classical channels. When the number of channel uses approaching infinity, the average error probability can be 0 if the transmission rate is smaller than capacity, and it becomes 1 if the transmission rate is larger than capacity [27, 28]. Recently, the exact relation how the average error probability depends on a smaller order of the transmission rate, in the case that the transmission rate (of the first order) equals the capacity, has been

derived independently in [25] and [26], which proves to be quite useful in evaluating the channel coding rate for finite blocklength case [25, 26].

There is a deep connection between hypothesis testing and channel capacity, both in the classical regime [29, 30] and in the quantum regime [31]. Recently, this connection has been generalized to the one-shot scenario as well [32, 33]. Indeed, such a connection is very helpful in the derivation of the second order coding rate in classical channel coding [25]. In contrast to its classical counterpart, which is an straightforward consequence of the central limit theorem, the second order asymptotics for quantum hypothesis testing is still unknown. In this paper, we solve this problem. Due to the above-mentioned connection, our result also make it possible to investigate the second order asymptotic behavior of classical information transmission over quantum channels [34].

Given a large number n of identical quantum systems, which are either of the state $\rho^{\otimes n}$ (the null hypothesis) or of the state $\sigma^{\otimes n}$ (the alternative hypothesis), we want to identify which state the systems belong to. Without loss of generality, this can be done by applying a two-outcome positive operator-valued measurement (POVM) $\{A_n, \mathbb{1} - A_n\}$, with $0 \leq A_n \leq \mathbb{1}$, on the joint Hilbert space $\mathcal{H}^{\otimes n}$ of the quantum systems. If we obtain the outcome associated to A_n , then we conclude that the state is $\rho^{\otimes n}$. Similarly, the outcome associated to $(\mathbb{1} - A_n)$ corresponds to the state $\sigma^{\otimes n}$. The error probabilities of the first kind and the second kind (also known as the type I and type II errors) are, respectively, defined as $\alpha_n(A_n) := \text{Tr}(\rho^{\otimes n}(\mathbb{1} - A_n))$ and $\beta_n(A_n) := \text{Tr}(\sigma^{\otimes n} A_n)$. Thus, $\alpha_n(A_n)$ is the probability that we incorrectly accept $\sigma^{\otimes n}$ while it is actually $\rho^{\otimes n}$, and $\beta_n(A_n)$ is the probability of the opposite situation.

Obviously, these two kinds of errors can be made arbitrarily small when n is big enough, unless $\rho = \sigma$, because $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ are asymptotically distinguishable. In an asymmetric setting, we want to maximize the exponential rate at which the type II error goes to 0, under the condition that the type I error simply converges to 0. Quantum Stein's lemma tells us that this maximal expo-

nent is the quantum relative entropy, which is given by $D(\rho\|\sigma) = \text{Tr}(\rho(\log \rho - \log \sigma))$ if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and $D(\rho\|\sigma) = \infty$ otherwise. It also says that if the type II error goes to 0 with an exponent bigger than $D(\rho\|\sigma)$, then the type I error must converge to 1. Formally, we have the statement as follows. (direct part [7]): For arbitrary $R < D(\rho\|\sigma)$, there exists a sequence of measurements $\{A_n, \mathbb{1} - A_n\}_n$, such that $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(A_n) \geq R$ and $\lim_{n \rightarrow \infty} \alpha_n(A_n) = 0$; (strong converse [8]): if a sequence of measurements $\{A_n, \mathbb{1} - A_n\}_n$ is such that $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(A_n) > D(\rho\|\sigma)$, then $\lim_{n \rightarrow \infty} \alpha_n(A_n) = 1$.

So, from the quantum Stein's lemma, we see that the error probability of the first kind jumps sharply from zero to one when the error exponent of the second kind passes by the relative entropy $D(\rho\|\sigma)$ from smaller to larger. One would ask how this sudden change happens at the point that the error exponent of the second kind is exactly $D(\rho\|\sigma)$. Intuitively, it should depend on a smaller order of the error exponent. In this paper, we have determined what this smaller order is, and how the error probability of the first kind depends on it. After defining the error-dependency function, we present our result in Theorem 2 as follows.

Definition 1 We define the function $\alpha_n(E_1, E_2|f)$, which reflects the dependency of the optimal error probability of the first kind on the error exponent of the second kind, up to the order n and \sqrt{n} , as

$$\alpha_n(E_1, E_2|f) := \min_{A_n} \left\{ \alpha_n(A_n) \mid \beta_n(A_n) \leq \exp \left(- (E_1 n + E_2 \sqrt{n} + f(n)) \right) \right\}, \quad (1)$$

where $f(n)$ is a function of some order other than n and \sqrt{n} , which is to be specified when necessary.

Theorem 2 Let $\Phi(x)$ be the cumulative distribution function of the standard normal distribution, i.e., $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$, we have

$$\lim_{n \rightarrow \infty} \alpha_n(E_1, E_2|f) = \begin{cases} 0 & \text{if } E_1 < D(\rho\|\sigma), f \in o(n) \\ \Phi \left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}} \right) & \text{if } E_1 = D(\rho\|\sigma), f \in o(\sqrt{n}) \\ 1 & \text{if } E_1 > D(\rho\|\sigma), f \in o(n), \end{cases} \quad (2)$$

where $V(\rho\|\sigma)$, which we name the quantum relative variance of ρ and σ , is defined as

$$V(\rho\|\sigma) := \text{Tr} \rho(\log \rho - \log \sigma)^2 - D^2(\rho\|\sigma).$$

Remark 1: The first and third cases of Eq. (2) in Theorem 2 are nothing else but the direct part and strong converse of quantum Stein's lemma, respectively. In fact, these two cases can be derived from the second case (see

the "Proof of Theorem 2" section). We include them in our theorem such that one easily gets the full information at first sight. Besides, our result of the first case is a bit stronger than the usual statement of the direct part of quantum Stein's lemma, because the former is equivalent to saying "for arbitrary $R < D(\rho\|\sigma)$ and arbitrary $\{\beta_n\}_n$ satisfying $\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n = R$, there exists a sequence of measurements $\{A_n, \mathbb{1} - A_n\}_n$, such that $\beta_n(A_n) \leq \bar{\beta}_n$ and $\lim_{n \rightarrow \infty} \alpha_n(A_n) = 0$ ".

Remark 2: If $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, we have $D(\rho\|\sigma) = +\infty$. Asymptotically, the optimal error probability of the first kind is always 0, while the error exponent of the second kind can be arbitrarily large. In such a case, the second order asymptotics and the strong converse make no sense. So, in this paper, we suppose $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and without loss of generality, we further suppose σ is of full rank.

Proof of Theorem 2. At first, we argue that it suffices to prove the second case, namely, for any $f(n) \in o(\sqrt{n})$,

$$\lim_{n \rightarrow \infty} \alpha_n(D(\rho\|\sigma), E_2|f) = \Phi \left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}} \right). \quad (3)$$

From Definition 1, it is easy to see that, for arbitrary $E_1 < D(\rho\|\sigma)$, $E_2 \in \mathbb{R}$, $E'_2 \in \mathbb{R}$, $f(n) \in o(n)$ and $f'(n) \in o(\sqrt{n})$, there exists $N \in \mathbb{N}$, such that

$$\alpha_n(E_1, E_2|f) \leq \alpha_n(D(\rho\|\sigma), E'_2|f') \quad (4)$$

holds for all $n \geq N$. Now, in Eq. (4), letting $n \rightarrow \infty$ at both sides and then letting $E'_2 \rightarrow -\infty$ at the right side, we get from Eq. (3)

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} \alpha_n(E_1, E_2|f) \\ &\leq \lim_{E'_2 \rightarrow -\infty} \lim_{n \rightarrow \infty} \alpha_n(D(\rho\|\sigma), E'_2|f') \\ &= \lim_{E'_2 \rightarrow -\infty} \Phi \left(\frac{E'_2}{\sqrt{V(\rho\|\sigma)}} \right) \\ &= 0, \end{aligned}$$

which is exactly the first case of Eq. (2). In a similar way, we can prove that the third case of Eq. (2) is also contained in the second case. Then, we show that Eq. (3) is a direct consequence of Theorem 3 presented below. By the definition of $\alpha_n(E_1, E_2|f)$, it is not difficult to see that, the achievability part of Theorem 3 implies

$$\limsup_{n \rightarrow \infty} \alpha_n(D(\rho\|\sigma), E_2|f) \leq \Phi \left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}} \right) \quad (5)$$

for any $f(n) \in o(\sqrt{n})$, and the optimality part of Theorem 3 implies

$$\liminf_{n \rightarrow \infty} \alpha_n(D(\rho\|\sigma), E_2|f) \geq \Phi \left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}} \right) \quad (6)$$

for any $f(n) \in o(\sqrt{n})$. Since Eq. (5) and Eq. (6) together, in turn, led to Eq. (3), all we need to do is proving Theorem 3. \square

Theorem 3 *For quantum hypothesis testing with the null hypothesis $\rho^{\otimes n}$ and the alternative hypothesis $\sigma^{\otimes n}$, and the error probabilities of the first and second kinds denoted as $\alpha_n(A_n)$ and $\beta_n(A_n)$, respectively, we have (Achievability): For any $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, there exists a sequence of measurements $\{A_n, \mathbb{I} - A_n\}_n$, such that*

$$\beta_n(A_n) \leq \exp \left\{ - \left(nD(\rho \parallel \sigma) + E_2\sqrt{n} + f(n) \right) \right\}, \quad (7)$$

$$\limsup_{n \rightarrow \infty} \alpha_n(A_n) \leq \Phi \left(\frac{E_2}{\sqrt{V(\rho \parallel \sigma)}} \right); \quad (8)$$

(Optimality): If there is a sequence of measurements $\{A_n, \mathbb{I} - A_n\}_n$ such that

$$\beta_n(A_n) \leq \exp \left\{ - \left(nD(\rho \parallel \sigma) + E_2\sqrt{n} + f(n) \right) \right\} \quad (9)$$

holds for given $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, then

$$\liminf_{n \rightarrow \infty} \alpha_n(A_n) \geq \Phi \left(\frac{E_2}{\sqrt{V(\rho \parallel \sigma)}} \right). \quad (10)$$

Before proving Theorem 3, we do some necessary preparation. Write $\rho = \sum_x \lambda(x) |a_x\rangle\langle a_x|$ and $\sigma = \sum_y \mu(y) |b_y\rangle\langle b_y|$, where $\{|a_x\rangle\}_x$ and $\{|b_y\rangle\}_y$, each being an orthonormal basis of the Hilbert space \mathcal{H} , are the eigenvectors of ρ and σ , respectively. $\lambda(x)$ and $\mu(y)$ are the corresponding eigenvalues, which satisfy $0 \leq \lambda(x) \leq 1$, $0 < \mu(y) \leq 1$ and $\sum_x \lambda(x) = \sum_y \mu(y) = 1$. Note that for the reason stated in remark 2, we assume $\mu(y) \neq 0$. Let x^n denote the sequence $x_1 x_2 \cdots x_n$ and y^n denote $y_1 y_2 \cdots y_n$. For n copies of the states ρ and σ , we can write

$$\rho^{\otimes n} = \sum_{x^n} \lambda^n(x^n) |a_{x^n}^n\rangle\langle a_{x^n}^n| \quad (11)$$

with $\lambda^n(x^n) = \prod_{i=1}^n \lambda(x_i)$ and $|a_{x^n}^n\rangle = |a_{x_1}\rangle \otimes |a_{x_2}\rangle \cdots |a_{x_n}\rangle$, and similarly,

$$\sigma^{\otimes n} = \sum_{y^n} \mu^n(y^n) |b_{y^n}^n\rangle\langle b_{y^n}^n| \quad (12)$$

with $\mu^n(y^n) = \prod_{i=1}^n \mu(y_i)$ and $|b_{y^n}^n\rangle = |b_{y_1}\rangle \otimes |b_{y_2}\rangle \cdots |b_{y_n}\rangle$, where the subscripts of x and y label which systems they belong to. The eigenvectors of ρ can be written as superpositions of the eigenvectors of σ , namely, the basis $\{|b_y\rangle\}_y$. That is, we write $|a_x\rangle = \sum_y \gamma_{xy} |b_y\rangle$, where $\gamma_{xy} = \langle b_y | a_x \rangle \in \mathbb{C}$ and $\sum_x |\gamma_{xy}|^2 = \sum_y |\gamma_{xy}|^2 = 1$. In such a way, we have

$$|a_{x^n}^n\rangle = \sum_{y^n} \gamma_{x^n y^n}^n |b_{y^n}^n\rangle, \quad \text{with } \gamma_{x^n y^n}^n = \prod_{i=1}^n \gamma_{x_i y_i}. \quad (13)$$

Define a random variable X , with alphabet $\{x\}_{x=1}^{|\mathcal{H}|}$ and probability distribution $P_X(x) = \lambda(x)$. Let Y be another random variable with alphabet $\{y\}_{y=1}^{|\mathcal{H}|}$, which depends on X with conditional distribution $P_{Y|X}(y|x) = |\gamma_{xy}|^2$. So, the joint distribution of (X, Y) is $P_{X,Y}(x, y) = \lambda(x) |\gamma_{xy}|^2$. Operationally, this is the probability of obtaining (x, y) , when we measure the quantum state ρ , sequentially in the bases $\{|a_x\rangle\}_x$ and $\{|b_y\rangle\}_y$. Let $(X^n, Y^n) := (X_1, Y_1)(X_2, Y_2) \cdots (X_n, Y_n)$ be a sequence of independent and identically distributed random variable pairs, and each (X_i, Y_i) has the same distribution as (X, Y) . Then

$$P_{X^n, Y^n}(x^n, y^n) = \prod_{i=1}^n \lambda(x_i) |\gamma_{x_i y_i}|^2 = \lambda^n(x^n) |\gamma_{x^n y^n}^n|^2. \quad (14)$$

As functions of X and Y , $\lambda(X)$ and $\mu(Y)$ are also random variables, and so are $\lambda^n(X^n)$ and $\mu^n(Y^n)$. The following lemma expresses the quantum relative entropy and quantum relative variance as statistical quantities of classical random variables.

Lemma 4 *We have*

$$D(\rho \parallel \sigma) = \mathbb{E}_{(X,Y)} \log \frac{\lambda(X)}{\mu(Y)}, \quad (15)$$

$$V(\rho \parallel \sigma) = \text{Var}_{(X,Y)} \log \frac{\lambda(X)}{\mu(Y)}. \quad (16)$$

Note again that we are only interested in the case that σ has full rank, so $\mu(Y) > 0$ (see remark 2). During the computation of the right sides of Eq. (15) and Eq. (16), if $\lambda(x) = 0$, we let $\lambda(x) \log \lambda(x) := \lim_{z \rightarrow 0} z \log z = 0$, and $\lambda(x) \log^2 \lambda(x) := \lim_{z \rightarrow 0} z \log^2 z = 0$.

Proof. This is done by direct calculation. For functions v and w , it is easy to check that

$$\text{Tr } v(\rho) = \sum_x v(\lambda(x)) = \sum_{xy} v(\lambda(x)) |\gamma_{xy}|^2,$$

and

$$\begin{aligned} & \text{Tr } v(\rho) w(\sigma) \\ &= \text{Tr} \left(\sum_x v(\lambda(x)) |a_x\rangle\langle a_x| \right) \left(\sum_y w(\mu(y)) |b_y\rangle\langle b_y| \right) \\ &= \sum_{xy} v(\lambda(x)) w(\mu(y)) |\gamma_{xy}|^2. \end{aligned}$$

Using these two equations with proper v and w at every step when needed, we get

$$\begin{aligned} & \text{Tr } \rho(\log \rho - \log \sigma) \\ &= \sum_{xy} (\lambda(x) \log \lambda(x) |\gamma_{xy}|^2 - \lambda(x) \log \mu(y) |\gamma_{xy}|^2) \\ &= \sum_{xy} P_{X,Y}(x, y) \log \frac{\lambda(x)}{\mu(y)} = \mathbb{E} \left(\log \frac{\lambda(X)}{\mu(Y)} \right), \end{aligned} \quad (17)$$

and

$$\begin{aligned}
& \text{Tr } \rho (\log \rho - \log \sigma)^2 \\
&= \text{Tr } \rho \log^2 \rho - 2 \text{Tr } (\rho \log \rho) \log \sigma + \text{Tr } \rho \log^2 \sigma \\
&= \sum_{xy} (\lambda(x) \log^2 \lambda(x) |\gamma_{xy}|^2 - 2 \lambda(x) \log \lambda(x) \\
&\quad \times \log \mu(y) |\gamma_{xy}|^2 + \lambda(x) \log^2 \mu(y) |\gamma_{xy}|^2) \\
&= \sum_{xy} P_{X,Y}(x, y) \left(\log \frac{\lambda(x)}{\mu(y)} \right)^2 = \mathbb{E} \left(\log \frac{\lambda(X)}{\mu(Y)} \right)^2.
\end{aligned} \tag{18}$$

Eq. (17) confirms Eq. (15), and Eq. (17) and Eq. (18) together let to Eq. (16), thus we finish the proof of Lemma 4. \square

Remark 3: In [9] and [13], similar classical expressions were found for quantum relative entropy and the quantum Chernoff distance, and that for the latter was used as a very powerful tool in proving the optimality part of quantum Chernoff bound.

Proof of Theorem 3: achievability. For any fixed $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, let

$$L_n := \exp \{ nD(\rho \parallel \sigma) + E_2 \sqrt{n} + f(n) \}.$$

Associated with every x^n , we define a projector $Q_{x^n}^n$ as

$$Q_{x^n}^n := \sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |b_{y^n}^n\rangle\langle b_{y^n}^n|.$$

Write $|\xi_{x^n}^n\rangle := Q_{x^n}^n |a_{x^n}^n\rangle$. Referring to Eq. (13), we have

$$|\xi_{x^n}^n\rangle = \sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \gamma_{x^n y^n}^n |b_{y^n}^n\rangle. \tag{19}$$

Let A_n be the projector onto the space S_n that is spanned by $\{|\xi_{x^n}^n\rangle\}_{x^n}$. We claim that the sequence of measurements $\{A_n, \mathbb{I} - A_n\}_n$ is what we needed: it satisfies Eq. (7) and Eq. (8).

Arrange all the values of x^n , in such a way that the eigenvalues of $\rho^{\otimes n}$, $\lambda^n(x^n)$'s, are in an increasing order. This gives an ordering to the vectors $\{|\xi_{x^n}^n\rangle\}_{x^n}$ as well. Let $g: \{i\}_{i=1}^{|\mathcal{H}|^n} \rightarrow \{x^n\}$ be the bijection mapping the position of x^n to x^n itself, i.e., x^n is at the $g^{-1}(x^n)$ th position in the above ordering. Then we have

$$\lambda^n(g(1)) \leq \lambda^n(g(2)) \leq \dots \leq \lambda^n(g(|\mathcal{H}|^n)). \tag{20}$$

Applying a modified Gram-Schmidt orthonormalization process to the sequence of vectors

$$|\xi_{g(1)}^n\rangle, |\xi_{g(2)}^n\rangle, |\xi_{g(3)}^n\rangle, \dots, |\xi_{g(|\mathcal{H}|^n)}^n\rangle,$$

we obtain a new sequence of vectors

$$|\hat{\xi}_{g(1)}^n\rangle, |\hat{\xi}_{g(2)}^n\rangle, |\hat{\xi}_{g(3)}^n\rangle, \dots, |\hat{\xi}_{g(|\mathcal{H}|^n)}^n\rangle. \tag{21}$$

The modification is that if $|\xi_{g(i)}^n\rangle \in \text{Span}(\{|\xi_{g(j)}^n\rangle\}_{j=1}^{i-1})$, (this including the case that $|\xi_{g(i)}^n\rangle = 0$), we let $|\hat{\xi}_{g(i)}^n\rangle =$

0. As a result, the set of vectors $\{|\hat{\xi}_{x^n}^n\rangle\}_{x^n}$ consists of an orthonormal basis of the space S_n , plus some zero vectors. Thus

$$A_n = \sum_{x^n} |\hat{\xi}_{x^n}^n\rangle\langle \hat{\xi}_{x^n}^n|. \tag{22}$$

The vectors $\{|\hat{\xi}_{x^n}^n\rangle\}_{x^n}$ have another property as follows. From the Gram-Schmidt process, we know that

$$|\hat{\xi}_{g(i)}^n\rangle = \sum_{j=1}^i s_{ij}^n |\xi_{g(j)}^n\rangle \tag{23}$$

for all $1 \leq i \leq |\mathcal{H}|^n$, with the coefficients $s_{ij}^n \in \mathbb{C}$. Further, from Eqs. (19, 20, 23), and paying attention to the definition of g , we conclude that

$$|\hat{\xi}_{x^n}^n\rangle = \sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} t_{x^n y^n}^n |b_{y^n}^n\rangle, \tag{24}$$

where $t_{x^n y^n}^n \in \mathbb{C}$ and

$$\sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |t_{x^n y^n}^n|^2 = 1. \tag{25}$$

Eqs. (12, 24, 25) led to

$$\begin{aligned}
& \text{Tr}(\sigma^{\otimes n} |\hat{\xi}_{x^n}^n\rangle\langle \hat{\xi}_{x^n}^n|) \\
&= \sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |t_{x^n y^n}^n|^2 \mu^n(y^n) \\
&\leq \frac{\lambda^n(x^n)}{L_n}.
\end{aligned} \tag{26}$$

So, making use of Eq. (22) and Eq. (26), we arrive at

$$\begin{aligned}
\beta_n(A_n) &= \text{Tr} \sigma^{\otimes n} A_n \\
&= \sum_{x^n} \text{Tr}(\sigma^{\otimes n} |\hat{\xi}_{x^n}^n\rangle\langle \hat{\xi}_{x^n}^n|) \\
&\leq \sum_{x^n} \frac{\lambda^n(x^n)}{L_n} \\
&= \frac{1}{L_n} = \exp \{ -(nD(\rho \parallel \sigma) + E_2 \sqrt{n} + f(n)) \},
\end{aligned}$$

which is Eq. (7).

On the other hand, Eq. (8) is confirmed as follows. Let

$$|\bar{\xi}_{x^n}^n\rangle := \begin{cases} 0 & \text{if } |\xi_{x^n}^n\rangle = 0 \\ \frac{|\xi_{x^n}^n\rangle}{\sqrt{\langle \xi_{x^n}^n | \xi_{x^n}^n \rangle}} & \text{if } |\xi_{x^n}^n\rangle \neq 0. \end{cases} \tag{27}$$

Obviously, $|\bar{\xi}_{x^n}^n\rangle \in S_n$. So

$$|\bar{\xi}_{x^n}^n\rangle\langle \bar{\xi}_{x^n}^n| \leq A_n. \tag{28}$$

Then we have

$$\begin{aligned}
\alpha_n(A_n) &= 1 - \text{Tr}(\rho^{\otimes n} A_n) \\
&\leq 1 - \sum_{x^n} \lambda^n(x^n) \text{Tr}(|a_{x^n}^n\rangle\langle a_{x^n}^n|(|\bar{\xi}_{x^n}^n\rangle\langle \bar{\xi}_{x^n}^n|)) \\
&= 1 - \sum_{x^n} \lambda^n(x^n) \langle \xi_{x^n}^n | \xi_{x^n}^n \rangle \\
&= 1 - \sum_{(x^n, y^n): \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \lambda^n(x^n) |\gamma_{x^n y^n}^n|^2 \\
&= 1 - \sum_{(x^n, y^n): \lambda^n(x^n)/\mu^n(y^n) \geq L_n} P_{X^n, Y^n}(x^n, y^n) \\
&= \text{Pr} \left\{ \frac{\lambda^n(X^n)}{\mu^n(Y^n)} < L_n \right\},
\end{aligned}$$

where the second line is by Eq. (11) and Eq. (28), the third line can be seen from the definition of $|\xi_{x^n}^n\rangle$ and $|\bar{\xi}_{x^n}^n\rangle$, the fourth line follows from Eq. (19) and the fifth line is due to Eq. (14). Noting that $\lambda^n(X^n) = \prod_{i=1}^n \lambda(X_i)$ and $\mu^n(Y^n) = \prod_{i=1}^n \mu(Y_i)$, and by taking logarithms at both sides, we see that $\frac{\lambda^n(X^n)}{\mu^n(Y^n)} < L_n$ is equivalent to

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho \| \sigma) \right) < E_2 + \frac{f(n)}{\sqrt{n}}.$$

As a result, we arrive at

$$\begin{aligned}
&\limsup_{n \rightarrow \infty} \alpha_n(A_n) \\
&\leq \lim_{n \rightarrow \infty} \text{Pr} \left\{ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho \| \sigma) \right) < E_2 + \frac{f(n)}{\sqrt{n}} \right\}
\end{aligned} \tag{29}$$

Since $f(n) \in o(\sqrt{n})$, we have for any $\epsilon > 0$, there exists $N(\epsilon)$, such that

$$E_2 + \frac{f(n)}{\sqrt{n}} < E_2 + \epsilon$$

holds for all $n \geq N(\epsilon)$. This means that, the right side of Eq. (29) is no larger than

$$\lim_{n \rightarrow \infty} \text{Pr} \left\{ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho \| \sigma) \right) < E_2 + \epsilon \right\},$$

which, by Lemma 4 and the central limit theorem, equals to $\Phi \left(\frac{E_2 + \epsilon}{\sqrt{V(\rho \| \sigma)}} \right)$. Hence,

$$\limsup_{n \rightarrow \infty} \alpha_n(A_n) \leq \Phi \left(\frac{E_2 + \epsilon}{\sqrt{V(\rho \| \sigma)}} \right).$$

Because ϵ is arbitrary, we conclude the proof of Eq. (8). \square

Proof of Theorem 3: optimality. Suppose that the sequence of measurements $\{A_n, \mathbb{1} - A_n\}_n$ satisfies Eq. (9). We will prove Eq. (10). Let

$$L_n := \exp \left\{ (nD(\rho \| \sigma) + E_2\sqrt{n} + f(n)) - f'(n) \right\} \tag{30}$$

with some fixed $f'(n) \in o(\sqrt{n})$ and $\lim_{n \rightarrow \infty} f'(n) = +\infty$. This ensures that

$$\lim_{n \rightarrow \infty} L_n \beta_n(A_n) = 0. \tag{31}$$

Associated with every x^n , we define the projector $Q_{x^n}^n$ as

$$Q_{x^n}^n := \sum_{y^n: \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |b_{y^n}^n\rangle\langle b_{y^n}^n|. \tag{32}$$

Inserting Eq. (11) into the definition of $\alpha_n(A_n)$, namely, $\alpha_n(A_n) := \text{Tr}(\rho^{\otimes n}(\mathbb{1} - A_n))$, and after a few calculations, we get

$$\alpha_n(A_n) = 1 - C_n - D_n, \tag{33}$$

where C_n and D_n are

$$C_n := \sum_{x^n} \lambda^n(x^n) \text{Tr} \left(Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\langle a_{x^n}^n| \sqrt{A_n} Q_{x^n}^n \right) \tag{34}$$

$$D_n := \sum_{x^n} \lambda^n(x^n) \text{Tr} \left((\mathbb{1} - Q_{x^n}^n) \sqrt{A_n} |a_{x^n}^n\rangle\langle a_{x^n}^n| \sqrt{A_n} (\mathbb{1} - Q_{x^n}^n) \right) \tag{35}$$

The basic difficulty in bounding C_n and D_n is that the POVM element A_n is very general except for the constraint of Eq. (9). Nevertheless, we will be able to show that the D_n term is asymptotically negligible, due to the constraint of Eq. (9) and our choice of L_n . This in turn, ensures that the C_n term can be upper bounded by removing the operator “ $\sqrt{A_n}$ ” from its expression, with only an infinitesimal correction (see Eq. (44)).

Now, we show that the D_n term is asymptotically negligible. Because

$$\begin{aligned}
\sigma^{\otimes n} &\geq (\mathbb{1} - Q_{x^n}^n) \sigma^{\otimes n} (\mathbb{1} - Q_{x^n}^n) \\
&\geq \frac{\lambda^n(x^n)}{L_n} (\mathbb{1} - Q_{x^n}^n),
\end{aligned}$$

where the first inequality is owing to the commutativity of $\sigma^{\otimes n}$ and the projector $(\mathbb{1} - Q_{x^n}^n)$, and the second one can be seen from the definition of $Q_{x^n}^n$, we obtain

$$\begin{aligned}
\beta_n(A_n) &= \text{Tr}(\sigma^{\otimes n} A_n) \\
&= \text{Tr} \left(\sigma^{\otimes n} \sqrt{A_n} \left(\sum_{x^n} |a_{x^n}^n\rangle\langle a_{x^n}^n| \right) \sqrt{A_n} \right) \\
&= \sum_{x^n} \text{Tr} \left(\sigma^{\otimes n} \left(\sqrt{A_n} |a_{x^n}^n\rangle\langle a_{x^n}^n| \sqrt{A_n} \right) \right) \\
&\geq \sum_{x^n} \text{Tr} \left(\frac{\lambda^n(x^n)}{L_n} (\mathbb{1} - Q_{x^n}^n) (\sqrt{A_n} |a_{x^n}^n\rangle\langle a_{x^n}^n| \sqrt{A_n}) \right) \\
&= \frac{D_n}{L_n}.
\end{aligned}$$

This result, which is rephrased as $D_n \leq L_n \beta_n(A_n)$, together with Eq. (31) tells us that

$$\lim_{n \rightarrow \infty} D_n = 0. \tag{36}$$

The evaluation of the C_n term will be a bit more complicated. For simplicity, we use the notation of norm, $\|\cdot\|$, defined as

$$\|\varphi\| := \sqrt{\langle \varphi | \varphi \rangle} = \sqrt{\text{Tr} |\varphi\rangle\langle \varphi|}, \quad (37)$$

for $|\varphi\rangle$ being a vector of some Hilbert space. Thus, C_n is rewritten as

$$C_n = \sum_{x^n} \lambda^n(x^n) \|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2. \quad (38)$$

Our strategy is to divide the terms in the sum of the above expression, into different classes, each satisfies some special conditions. Then we evaluate them individually under these conditions. For such a purpose, we define index sets

$$\begin{aligned} \mathcal{O}_1^n &:= \left\{ x^n \mid \|\sqrt{A_n} |a_{x^n}^n\rangle\| \geq \epsilon_1 \right\}, \\ \mathcal{O}_2^n &:= \left\{ x^n \mid \|(\mathbb{1} - Q_{x^n}^n) \sqrt{A_n} |a_{x^n}^n\rangle\| \leq \epsilon_1 \epsilon_2 \right\} \end{aligned}$$

with sufficiently small $\epsilon_1, \epsilon_2 > 0$. Denote the full set of all the x^n 's as \mathcal{O}^n , and the complementary sets of \mathcal{O}_1^n and \mathcal{O}_2^n as $\overline{\mathcal{O}_1^n}$ and $\overline{\mathcal{O}_2^n}$, respectively. Since \mathcal{O}^n is the union of three disjoint subsets

$$\mathcal{O}^n = \overline{\mathcal{O}_1^n} \cup (\mathcal{O}_1^n \cap \overline{\mathcal{O}_2^n}) \cup (\mathcal{O}_1^n \cap \mathcal{O}_2^n),$$

we deal with Eq. (38) under distinct cases that x^n belongs to these subsets respectively, and then sum them up.

The first case is that $x^n \in \overline{\mathcal{O}_1^n}$. Noting that a projector (more generally, any contraction operator whose singular values are no more than 1) acting on a vector will only reduce its norm, we have

$$\begin{aligned} & \sum_{x^n \in \overline{\mathcal{O}_1^n}} \lambda^n(x^n) \|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & \leq \sum_{x^n \in \overline{\mathcal{O}_1^n}} \lambda^n(x^n) \epsilon_1^2 \leq \sum_{x^n} \lambda^n(x^n) \epsilon_1^2 \\ & = \epsilon_1^2. \end{aligned} \quad (39)$$

The second case is that $x^n \in \mathcal{O}_1^n \cap \overline{\mathcal{O}_2^n}$. We upper bound it as

$$\begin{aligned} & \sum_{x^n \in \mathcal{O}_1^n \cap \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & \leq \sum_{x^n \in \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \\ & \leq \sum_{x^n \in \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \frac{1}{\epsilon_1^2 \epsilon_2^2} \|(\mathbb{1} - Q_{x^n}^n) \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & \leq \frac{1}{\epsilon_1^2 \epsilon_2^2} \sum_{x^n} \lambda^n(x^n) \|(\mathbb{1} - Q_{x^n}^n) \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & = \frac{D_n}{\epsilon_1^2 \epsilon_2^2}, \end{aligned} \quad (40)$$

where the second line follows from inequalities $\|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\| \leq \|\sqrt{A_n} |a_{x^n}^n\rangle\| \leq \| |a_{x^n}^n\rangle \| = 1$ and enlarging the range of x^n in the sum, the third line is by the definition of \mathcal{O}_2^n , in the fourth line, we enlarge the range of x^n further, and for the last line, it can be easily seen from Eq. (35) and Eq. (37). The last case, which will turn out to be the dominant part, is that $x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n$. In such a case, paying attention to the definition of \mathcal{O}_1^n and \mathcal{O}_2^n , we see that

$$\begin{aligned} & \left\| \frac{\sqrt{A_n} |a_{x^n}^n\rangle}{\|\sqrt{A_n} |a_{x^n}^n\rangle\|} - Q_{x^n}^n \frac{\sqrt{A_n} |a_{x^n}^n\rangle}{\|\sqrt{A_n} |a_{x^n}^n\rangle\|} \right\| \\ & = \frac{\|(\mathbb{1} - Q_{x^n}^n) \sqrt{A_n} |a_{x^n}^n\rangle\|}{\|\sqrt{A_n} |a_{x^n}^n\rangle\|} \leq \frac{\epsilon_1 \epsilon_2}{\epsilon_1} = \epsilon_2. \end{aligned}$$

Then, directly applying Lemma 5, which will be presented and proven later, we get

$$\begin{aligned} & \left\| \frac{\sqrt{A_n} |a_{x^n}^n\rangle \langle a_{x^n}^n| \sqrt{A_n} |a_{x^n}^n\rangle}{\|\sqrt{A_n} |a_{x^n}^n\rangle\|^2} \right\|^2 \\ & \leq \left\| \left(Q_{x^n}^n \frac{\sqrt{A_n} |a_{x^n}^n\rangle \langle a_{x^n}^n| \sqrt{A_n} |a_{x^n}^n\rangle}{\|\sqrt{A_n} |a_{x^n}^n\rangle\|^2} Q_{x^n}^n \right) |a_{x^n}^n\rangle \right\|^2 + 2\sqrt{2}\epsilon_2. \end{aligned} \quad (41)$$

Since $0 \leq A_n \leq \mathbb{1}$, it holds that $A_n \leq \sqrt{A_n}$. As a result,

$$\begin{aligned} & \|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \leq \|\sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & \leq \left\| \frac{\langle a_{x^n}^n | \sqrt{A_n} |a_{x^n}^n\rangle}{\langle a_{x^n}^n | A_n |a_{x^n}^n\rangle} \sqrt{A_n} |a_{x^n}^n\rangle \right\|^2. \end{aligned} \quad (42)$$

The last term of Eq. (42) and the left side of Eq. (41) are actually the same. So, combining these two equations together, and noting that the right side of Eq. (41) is obviously upper bounded by

$$\|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2,$$

we arrive at

$$\|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \leq \|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2.$$

Thus,

$$\begin{aligned} & \sum_{x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n} \lambda^n(x^n) \|Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ & \leq \sum_{x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n} \lambda^n(x^n) (\|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2) \\ & \leq \sum_{x^n} \lambda^n(x^n) \|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2. \end{aligned} \quad (43)$$

Now, adding Eq. (39), Eq. (40) and Eq. (43) together, we obtain from Eq. (38) that

$$C_n \leq \sum_{x^n} \lambda^n(x^n) \|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 + \frac{D_n}{\epsilon_1^2 \epsilon_2^2} + \epsilon_1^2 + 2\sqrt{2}\epsilon_2. \quad (44)$$

In analogy to what we did during the derivation of Eq. (8) in the proof of the achievability part of Theorem 3, the limit of the first term of the right side of Eq. (44) is computed to be

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \sum_{x^n} \lambda^n(x^n) \|Q_{x^n}^n |a_{x^n}^n\rangle\|^2 \\
&= \lim_{n \rightarrow \infty} \sum_{(x^n, y^n): \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \lambda^n(x^n) |\gamma_{x^n y^n}^n|^2 \\
&= \lim_{n \rightarrow \infty} \Pr_{(X^n, Y^n)} \left\{ \frac{\lambda^n(X^n)}{\mu^n(Y^n)} \geq L_n \right\} \\
&= \lim_{n \rightarrow \infty} \Pr_{(X^n, Y^n)} \left\{ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho \parallel \sigma) \right) \right. \\
&\quad \left. \geq E_2 + \frac{1}{\sqrt{n}} (f(n) - f'(n)) \right\} \\
&= 1 - \Phi \left(\frac{E_2}{\sqrt{V(\rho \parallel \sigma)}} \right),
\end{aligned} \tag{45}$$

where the first equality can be seen from Eq. (13) and Eq. (32), the second and third equalities make use of Eq. (14) and Eq. (30), respectively, the last equality follows from Lemma 4 and the central limit theorem, as well as the fact that $f(n), f'(n) \in o(\sqrt{n})$. Hence, due to Eq. (36) and Eq. (45), and then noting that $\varepsilon_1, \varepsilon_2 > 0$ can be arbitrarily small, we get from Eq. (44) that

$$\limsup_{n \rightarrow \infty} C_n \leq 1 - \Phi \left(\frac{E_2}{\sqrt{V(\rho \parallel \sigma)}} \right). \tag{46}$$

Eventually, inserting Eq. (36) and Eq. (46) into Eq. (33) results in

$$\liminf_{n \rightarrow \infty} \alpha_n(A_n) \geq \Phi \left(\frac{E_2}{\sqrt{V(\rho \parallel \sigma)}} \right)$$

which is Eq. (10) and we finish the proof of the optimality part of Theorem 3. \square

Lemma 5 *Let $|\phi\rangle$ and $|\varphi\rangle$ be normalized vectors in some Hilbert space. Let π be a projector. if $\| |\phi\rangle - \pi|\phi\rangle \| \leq \varepsilon$, then*

$$\| (|\phi\rangle\langle\phi|)|\varphi\rangle \|^2 - \| (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \|^2 \leq 2\sqrt{2}\varepsilon. \tag{47}$$

Proof. We show Eq. (47) as follows.

$$\begin{aligned}
& \| (|\phi\rangle\langle\phi|)|\varphi\rangle \|^2 - \| (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \|^2 \\
&= (\| (|\phi\rangle\langle\phi|)|\varphi\rangle \| + \| (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \|) \\
&\quad \times (\| (|\phi\rangle\langle\phi|)|\varphi\rangle \| - \| (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \|) \\
&\leq 2(\| (|\phi\rangle\langle\phi|)|\varphi\rangle \| - \| (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \|) \\
&\leq 2\| (|\phi\rangle\langle\phi|)|\varphi\rangle - (\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle \| \\
&= 2\| (\langle\phi|(\mathbb{1} - \pi)|\varphi\rangle)\pi|\phi\rangle + (\langle\phi|\varphi\rangle)(\mathbb{1} - \pi)|\phi\rangle \| \\
&= 2\sqrt{|\langle\phi|(\mathbb{1} - \pi)|\varphi\rangle|^2 \cdot \|\pi|\phi\rangle\|^2 + |\langle\phi|\varphi\rangle|^2 \cdot \|(\mathbb{1} - \pi)|\phi\rangle\|^2} \\
&\leq 2\sqrt{\| |\phi\rangle - \pi|\phi\rangle \|^2 \cdot 1 + 1 \cdot \| |\phi\rangle - \pi|\phi\rangle \|^2} \\
&\leq 2\sqrt{\varepsilon^2 + \varepsilon^2} = 2\sqrt{2}\varepsilon,
\end{aligned}$$

where the fifth line is by the triangle inequality, the seventh line is due to the Pythagoras' theorem, and the other lines are trivially by direct calculations and the conditions stated in the lemma. \square

Conclusion. we have obtained the second order asymptotics for quantum hypothesis testing. Our result deepens the quantum Stein's lemma, in a way just like the central limit theorem does to the law of large numbers. Our method is elementary, based on basic linear algebra and probability theory. It deals with the achievability part and the converse part in a unified framework, with a clear geometric picture. We hope that the new result obtained and the new method employed, in this paper, will find important applications in other aspects of quantum information and, non-commutative probability and statistics.

Acknowledgments. The author would like to thank Dr. William Matthews for introducing the results of [25] and [26], and Prof. Shunlong Luo for a conversation on related topics. He is especially grateful to Prof. Masahito Hayashi for many helpful discussions. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

* Electronic address: carl.ke.lee@gmail.com

- [1] T. M. Cover, J. A. Thomas, *Elements of Information Theory* (New York: Wiley, 1991).
- [2] H. Chernoff, Ann. Math. Stat. **23**, 493 (1952).
- [3] W. Hoeffding, Ann. Math. Stat. **36**, 369 (1965).
- [4] I. Csiszár, G. Longo, Studia Sci. Math. Hungarica **6**, 181 (1971).
- [5] R. E. Blahut, IEEE Trans. Inf. Theory **20**, 405 (1974).
- [6] T. S. Han, K. Kobayashi, IEEE Trans. Inf. Theory **35**, 178 (1989).
- [7] F. Hiai, D. Petz, Comm. Math. Phys. **143**, 99 (1991).
- [8] T. Ogawa, H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).

- [9] M. Nussbaum, A. Szkoła, *Ann. Stat.* **37**, 1040 (2009)
- [10] K. M. R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Muñoz-Tapia, A. Acín, E. Bagan, F. Verstraete, *Phys. Rev. Lett.* **98**, 160501 (2007).
- [11] M. Hayashi, *Phys. Rev. A* **76**, 062301 (2007).
- [12] H. Nagaoka, Arxiv preprint quant-ph/0611289 (2006).
- [13] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, F. Verstraete, *Comm. Math. Phys.* **279**, 251 (2008).
- [14] M. Hayashi, *J. Phys. A: Math. Gen.* **35** 10759 (2002).
- [15] M. Hayashi, *Quantum Information, An Introduction* (Berlin: Springer, 2006).
- [16] H. Nagaoka, M. Hayashi, *IEEE Trans. Inf. Theory* **53**, 534 (2007).
- [17] I. Bjelaković, J. D. Deuschel, T. Krüger, R. Seiler, Ra. Siegmund-Schultze, A. Szola, *Comm. Math. Phys.* **260**, 659 (2005).
- [18] I. Bjelaković, J. D. Deuschel, T. Krüger, R. Seiler, Ra. Siegmund-Schultze, A. Szola, *Comm. Math. Phys.* **279**, 559 (2008).
- [19] F. G. S. L. Brandão, M. B. Plenio, *Comm. Math. Phys.* **295**, 791 (2010).
- [20] R. E. Kass, P. W. Vos, *Geometrical Foundations of Asymptotic Inference* (New York: Wiley, 1997).
- [21] V. Strassen, in *Proc. 3rd Conf. Inf. Theory*, Prague, Czech Republic, 1962, pp. 689-723.
- [22] I. Kontoyiannis, *IEEE Trans. Inf. Theory* **43**, 1339 (1997).
- [23] E. Figueroa, C. Houdre, *IEEE Trans. Inf. Theory* **51**, 1339 (2005).
- [24] M. Hayashi, *IEEE Trans. Inf. Theory* **54**, 4619 (2008).
- [25] M. Hayashi, *IEEE Trans. Inf. Theory* **55**, 4947 (2009).
- [26] Y. Polyanskiy, H. V. Poor, S. Verdú, *IEEE Trans. Inf. Theory* **56**, 2307 (2010).
- [27] C. E. Shannon, *Bell Syst. Tech. J.* **27**:379 (1948).
- [28] J. Wolfowitz, *Coding Theorems of Information Theory* (Berlin: Springer, 1964).
- [29] S. Verdú, T. S. Han, *IEEE Trans. Inf. Theory* **40**, 1147 (1994).
- [30] T. S. Han, *Information-Spectrum Methods in Information Theory* (Berlin: Springer, 2003).
- [31] M. Hayashi, H. Nagaoka, *IEEE Trans. Inf. Theory* **49**, 1753 (2003).
- [32] M. Mosonyi, N. Datta, *J. Math. Phys.* **50**, 072104 (2009).
- [33] Ligong Wang, R. Renner, *Phys. Rev. Lett.* **108**, 200501 (2012).
- [34] M. Hayashi, Ke Li, in preparation (2012).